DOCUMENT RESUME

ED 189 117 TM 800 281

AUTHOR Gialluca, Kathleen A.: Weiss, David J.
TITLE Effects of Immediate Knowledge of Results on Achievement Test Perfcrmance and Test Dimensionality. Research Report 80-1.
INSTITUTION Minnesota Univ., Minneapolis. Dept. of Psychology.
SPONS AGENCY Office of Naval Research, Arlington, Va. Personnel and Training Research Programs Office.
PUB DATE Jan 80
CONTRACT N00014-76-C-0627
NOTE 32p.

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS *Achievement Tests: Biology: *Computer Assisted Testing: Cues: Factor Structure: *Feedback: Higher Education: *Latent Trait Theory: Multidimensional Scaling: Science Tests: *Scores: Test Items
IDENTIFIERS Adaptive Testing

ABSTRACT

Two studies investigated the effects of immediate knowledge of results (feedback) of college biology test items upon performance on subsequent items. Students received correct answers after responding to each item. Tests and correct answers were administered on a cathode-ray terminal connected to a computer. Feedback did not systematically increase total test scores, as would be expected if students were using results from previously administered items as clues for subsequent items. Furthermore, feedback did not alter the latent trait and item characteristic curve assumption of unidimensionality, or local independence, among achievement test items. (CP)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED189117

MAY 1980
RECEIVED
ERIC/TM

EFFECTS OF IMMEDIATE KNOWLEDGE OF RESULTS ON ACHIEVEMENT TEST PERFORMANCE AND TEST DIMENSIONALITY

SCOPE OF INTEREST NOTICE
The ERIC Facility has assigned this document for processing to:
TM | IR
In our judgement, this document is also of interest to the clearinghouses noted to the right. Indexing should reflect their special points of view.

Kathleen A. Gialluca
and
David J. Weiss

RESEARCH REPORT 80-1
JANUARY 1980

COMPUTERIZED ADAPTIVE TESTING LABORATORY
PSYCHOMETRIC METHODS PROGRAM
DEPARTMENT OF PSYCHOLOGY
UNIVERSITY OF MINNESOTA
MINNEAPOLIS, MN 55455

This research was supported by funds from the Air Force Human Resources Laboratory, Defense Advanced Research Projects Agency, Navy Personnel Research and Development Center, Army Research Institute, and Office of Naval Research, and monitored by the Office of Naval Research.

Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

TM 800281

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER Research Report 80-1	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Effects of Immediate Knowledge of Results on Achievement Test Performance and Test Dimensionality		5. TYPE OF REPORT & PERIOD COVERED Technical Report 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Kathleen A. Gialluca and David J. Weiss		8. CONTRACT OR GRANT NUMBER(s) N00014-76-C-0627
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Psychology University of Minnesota Minneapolis, MN 55455		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS P.E.: 61153N PROJ.:RR042-04 T.A.: RR042-04-01 W.U.: NR150-389
11. CONTROLLING OFFICE NAME AND ADDRESS Personnel and Training Research Programs Office of Naval Research Arlington, VA 22217		12. REPORT DATE January 1980 13. NUMBER OF PAGES 25
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release; distribution unlimited. Reproduction in whole or in part is permitted for any purpose of the United States Government.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

This research was supported by funds from the Air Force Human Resources Laboratory, Defense Advanced Research Projects Agency, Navy Personnel Research and Development Center, Army Research Institute, and Office of Naval Research, and monitored by the Office of Naval Research.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Feedback	Adaptive Testing	Response-Contingent Testing
Testing	Tailored Testing	Latent Trait Theory
Achievement Testing	Knowledge of Results	Item Response Theory
Computerized Testing	Branched Testing	Item Characteristic Curve Theory

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

These two studies investigated the effects of administering immediate knowledge of results (KR) concerning the correctness or incorrectness of item response on a computerized adaptive test of biology achievement. In the case of incorrect responses, the correct answers were provided to the student.

The results of these studies indicate that the provision of informative KR did not systematically increase total test scores, as would be expected if

students were using information from previously administered items to help them answer subsequent items. Futhermore, provision of informative KR did not alter the dimensionality of the achievement tests administered, indicating that the latent trait model assumption of local independence among the items was not affected to any significant degree.

CONTENTS

Introduction 1

Study I 1

Method 1
Subjects 1
Procedure 2
Adaptive Test 2
Item Pools 2
Entry Level 3
Branching Strategy 3
Termination Criterion 3
Scoring 3
Factor Analyses 3

Results 4
Subtest Scores 4
Subtest Intercorrelations 6
Factor Analyses 7
Exam 1 7
Exam 2 9

Conclusions 11
Subtest Scores 11
Subtest Intercorrelations and Factor Analyses 11

Study II 12

Method 12
Subjects 12
Procedure 12
Adaptive Test 12

Results 13
Subtest Scores 13
Subtest Intercorrelations 13
Factor Analyses 15
Exam 1 15
Exam 2 17

Conclusions 19
Subtest Scores 19
Subtest Intercorrelations and Factor Analyses 19

Discussion and Conclusions 20

Subtest Scores 20
Factor Analyses 20
Further Research 21

References 22

Appendix: Supplementary Tables 24

MAY 1 1980

Acknowledgments

Data utilized in this report were obtained from volunteer students in General Biology, Biology 1-011, at the University of Minnesota, during fall quarter 1976, winter quarter 1977, and winter quarter 1978. Appreciation is extended to these students for their participation in this research. The cooperation of Kathy Swart and Norman Kerr of the General Biology staff in providing access to the students is also deeply appreciated.

Technical Editor: Barbara Leslie Camm

Effects of Immediate Knowledge of Results On Achievement Test Performance And Test Dimensionality

It is commonly accepted that sufficient levels of motivation during testing are necessary in order to adequately measure an individual's levels of ability and achievement (e.g., Cronbach, 1970; Samuda, 1975; Terman, 1916). Motivation may be enhanced or maintained by the provision of feedback or knowledge of results (KR) immediately after each item response (Betz, 1975; Betz & Weiss, 1976a, b). Although providing KR may be rather cumbersome when tests are administered in paper-and-pencil format, it is a rather trivial task when tests are administered by an interactive computer; it is within the context of computerized testing that the provision of KR may be most fruitfully employed.

KR may be of different forms. That is, students may be told only whether their responses to the items are correct or incorrect. Often, however, they may also be provided with the correct answer for those items answered incorrectly. Particularly in the context of achievement testing, this distinction may be important. If the correct answers to some test items contain clues or information regarding the correct answers to later items, then the provision of such informative KR may affect students' test scores to some unknown degree. The effect may well be cumulative in nature (i.e., greater toward the end of the test after most items have been administered) and may be quite distinct from the effects due to motivation alone. The introduction of this interdependence among achievement test items may affect the dimensionality of the achievement test to a certain degree. That is, other factors, unrelated to achievement per se, may come into play and account for some portion of test performance.

If it was originally assumed that item responses on an achievement test could be fit by an item characteristic curve (ICC) model (e.g., Birnbaum, 1968), then the provision of informative KR becomes problematic. A basic assumption of most ICC test models is local independence among the test items and, hence, unidimensionality of the test. To the extent that this assumption is violated, use of ICC models for tests that provide KR may be inappropriate.

The present report includes two studies that were designed to investigate the question of the appropriateness of ICC models under conditions of informative KR in an achievement testing environment.

STUDY 1

Method

Subjects

Data were obtained from students enrolled in an introductory biology course at the University of Minnesota during two academic quarters, fall 1976 and winter 1977. During each of these quarters,

volunteers were recruited to take computerized adaptive versions of their two classroom midquarter exams. Students received one point, which was added to their final course grade, for participating in one of the adaptive testing sessions and an additional two points for participating in both sessions. During the fall quarter, data were available for 377 students for the first computerized adaptive midquarter exam (Exam 1) and for 386 students for the second computerized adaptive midquarter exam (Exam 2). During winter quarter these figures were 317 and 349, respectively.

Procedure

For both Exam 1 and Exam 2, within each of the two quarters, the volunteer students were given three tests in the following order: (1) an adaptive verbal ability test, (2) the multiple-choice adaptive biology test, and (3) a test consisting of specially designed biology items. For the present report, only the data from the multiple-choice adaptive biology tests were analyzed.

Tests were administered on cathode-ray terminals (CRTs) connected to a real-time computer. Items were displayed on the CRT at 960 characters per second. Tests were preceded by a series of instructional screens (DeWitt & Weiss, 1974) designed to explain the operation of the equipment to the students.

During the fall quarter administration, KR was provided on the CRT screen after a student responded to each item, i.e., each student was informed whether or not he/she had answered each item correctly and was told the correct response to each item answered incorrectly. During the winter quarter administration, immediate KR was not provided. There were no time limits imposed on any of the three tests. At the completion of testing, students received a printed report that listed items answered incorrectly along with the correct answers.

Adaptive Test

Item pools. The development of the item pools used in this study has been described by Bejar, Weiss, and Kingsbury, 1977. The answer sheets for two classroom midquarter exams from two previous academic quarters were used as raw data for estimating the three ICC item parameters for each test item--discrimination (a), difficulty (b), and lower asymptote (c). For the fall quarter administration 114 items, which covered the content of the first classroom exam, were available for Exam 1. The pool for Exam 2 contained 112 items: This pool resulted from the removal of 11 items that had been administered in a special format from the 123-item pool used by Bejar et al. (1977). For the winter administration 44 items were added to the pool for Exam 1, and 49 were added to the pool for Exam 2. There were thus a total of 158 items in the Exam 1 item pool and 161 in the Exam 2 item pool.

To construct item pools that could be used for administration of stradaptive tests (Vale & Weiss, 1975a,b; Weiss, 1973), each of the two pools was structured by forming nine strata of increasing difficulty. Mean stratum difficulties were chosen so that there would be approximately the same number of items per stratum. Within each

stratum the items were ordered in terms of their discriminations unless this resulted in items covering the same content area appearing consecutively. Appendix Tables A and B show the item parameters for the items in the nine strata into which the Exam 1 and Exam 2 item pools were structured.

Entry level. One of the advantages of the stradaptive testing strategy is that prior information can be used to select the stratum from which the first item is administered. In this study the entry stratum was selected by the student; at the beginning of each stradaptive test, students were asked to state their grade-point average (GPA) by selecting one of nine equally spaced GPA intervals from 2.00 to 4.00 (DeWitt & Weiss, 1974, p. 48). On the assumption that overall GPA was related to biology achievement levels, students with the highest GPAs began the stradaptive test with an item at the most difficult stratum (Stratum 9), and those with the lowest GPA began with an item at the least difficult stratum (Stratum 1).

Branching strategy. The branching strategy used in the stradaptive test was (1) if the current item was answered incorrectly, or skipped, the next unadministered item from the next easier stratum was administered or (2) if the current item was answered correctly, the next unadministered item from the next more difficult stratum was administered.

Termination criterion. A variable criterion was used to terminate testing on the stradaptive test. After a student answered five items in a stratum, if he/she answered 20% or fewer correctly, testing was terminated. If testing was not terminated by this criterion after 50 items had been administered, no additional items were administered.

Scoring

Each of the computerized adaptive midquarter exams was scored in the following manner. For each student the exam was divided into four subtests of equal length; whenever the original test length was not evenly divisible by four, the first one, two, or three items were omitted and the analysis proceeded on the basis of the remaining items.

Owen's (1975) Bayesian procedure for scoring items was used separately for each of the subtests within each of the exams. The prior distribution of achievement level was assumed to be normal, with a mean of 0 and a standard deviation of 1 at the beginning of each of the subtests; this achievement level estimate ($\hat{\theta}$) was updated after each item. For scoring purposes, then, the four subtests were treated as independent units, and four achievement level estimates were therefore available for each student on each exam. This procedure allowed for an investigation of the existence of any effects of KR that might be cumulative in nature.

Factor Analyses

If the provision of immediate KR during the adaptive midquarter exams were to affect the dimensionality of test performance (i.e., by

violating the assumption of local independence among the items), then the factor structures of the exams with KR, administered during fall quarter, should differ somewhat from the factor structures of the exams without KR, administered during winter quarter.

Because these exams were adaptive, students necessarily received different subsets of items. Hence, analyses were not performed at the item level but, rather, at the level of subtest scores. That is, principal axes factor analyses were performed separately on the inter-correlations of the four scores from the intra-individually determined subtests of each midquarter exam. The diagonal elements of the inter-subtest correlation matrices were replaced with initial estimates of the communalities of each variable, as given by the squared multiple correlation between that variable and the other variables in the matrix. An iterative procedure for improving these communality estimates was used, successively extracting factors and re-estimating the communalities. This process continued until the differences between the two successive communality estimates were negligible.

Results

Subtest Scores

The means and standard deviations of the subtest achievement level estimates for each of the two computerized adaptive midquarter exams are given in Table 1 for each of the two quarters. The means

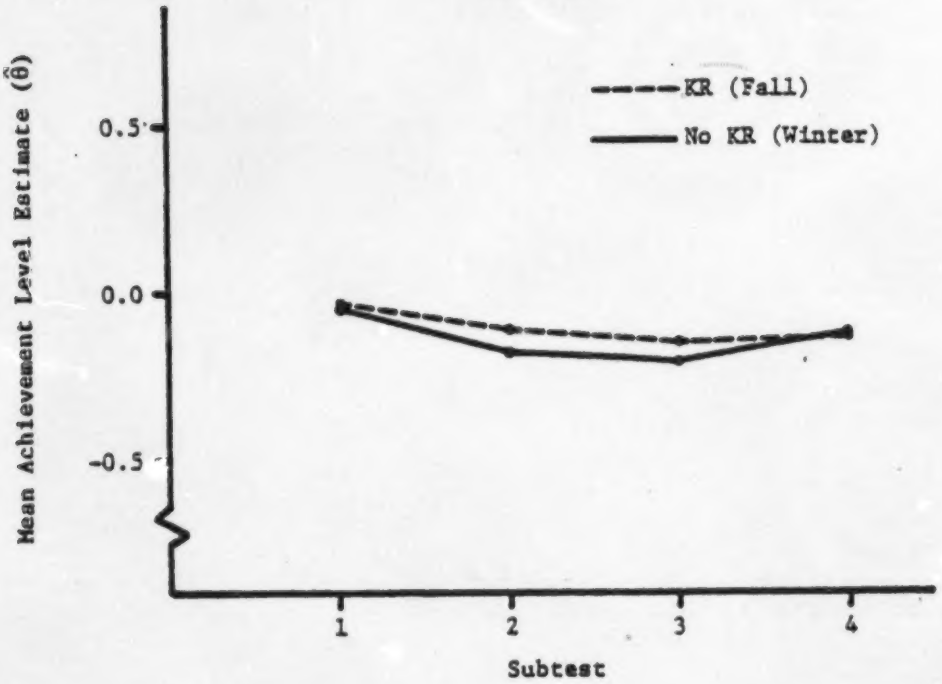
Table 1
Means and Standard Deviations (S.D.) of Subtest Achievement Level Estimates for Exams 1 and 2, for KR (Fall) and No-KR (Winter) Groups

Exam and Subtest	KR (Fall) Mean	KR (Fall) S.D.	No-KR (Winter) Mean	No-KR (Winter) S.D.
Exam 1				
Subtest 1	-.03	.86	-.04	.90
Subtest 2	-.11	.86	-.18	.83
Subtest 3	-.15	.82	-.21	.83
Subtest 4	-.14	.71	-.13	.73
Exam 2				
Subtest 1	-.11	.72	.00	.79
Subtest 2	-.18	.80	-.07	.73
Subtest 3	-.22	.79	-.11	.71
Subtest 4	-.18	.81	-.03	.70

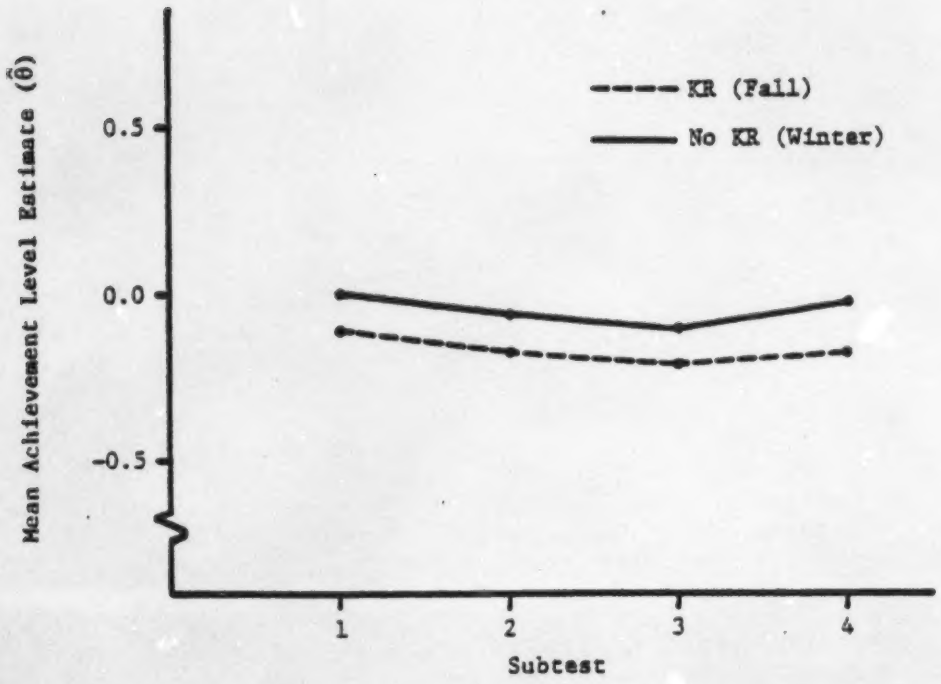
of these achievement level estimates ($\hat{\theta}$) are plotted, separately for each midquarter, in Figure 1.

Figure 1
Mean Achievement Level Estimates ($\hat{\theta}$) for Each Subtest

(a) Exam 1



(b) Exam 2



Since students in the fall quarter received immediate KR concerning their performance after responding to each item, and students in the winter quarter received no such KR, comparison of the subtest scores across academic quarters allowed for a comparison of the effects of the administration of KR on achievement level estimates.

Inspection of Table 1 and Figure 1 shows that there were no consistent differences between the two groups in the achievement level estimates obtained for Exam 1 and Exam 2. For Exam 1 (Figure 1a), the two curves are essentially identical; for Exam 2 (Figure 1b), there is a slightly larger discrepancy between the curves (favoring the no-KR group), although this discrepancy was very small (approximately .1 unit on the θ metric). None of the differences between mean achievement level estimates for the KR and no-KR groups on any of the subtests were statistically significant($p=.05$).

Subtest Intercorrelations

The inter-subtest correlation matrices for both exams in the KR and no-KR groups are given in Tables 2 and 3. Table 2 presents these correlations from Exam 1 subtests for the KR group (above diagonal) and the no-KR group (below diagonal). The correlations within each group were fairly homogeneous, ranging from .44 to .52 for the KR group and from .53 to .59 for the no-KR group. For Exam 1 the inter-subtest correlations for the no-KR group were uniformly higher than the inter-subtest correlations for the KR group.

Table 2
Exam 1 Subtest Intercorrelation Matrix for KR (Fall) Group (Above Diagonal) and No-KR (Winter) Group (Below Diagonal)

	Subtest			
Subtest	1	2	3	4
1		.44	.46	.49
2	.54		.51	.52
3	.53	.54		.47
4	.56	.57	.59	

Table 3 presents the inter-subtest correlations from the two academic quarters for Exam 2. Here, the within-quarter correlations were less homogeneous, ranging from .50 to .69 for the KR group and from .54 to .60 for the no-KR group. Unlike the Exam 1 results, where the no-KR correlations were uniformly higher than the KR correlations, there were no mean differences in the sizes of the correlations between groups. However, the KR correlations were much more heterogeneous than the no-KR correlations. The generally low levels of correlation among the four subtests in both groups were due to the fact that there were only an average of about seven items in each subtest.

Table 3
Exam 2 Subtest Intercorrelation Matrix for KR (Fall) Group (Above Diagonal) and No-KR (Winter) Group (Below Diagonal)

	Subtest			
Subtest	1	2	3	4
1		.50	.61	.64
2	.57		.58	.63
3	.54	.56		.69
4	.57	.59	.60	

Factor Analyses

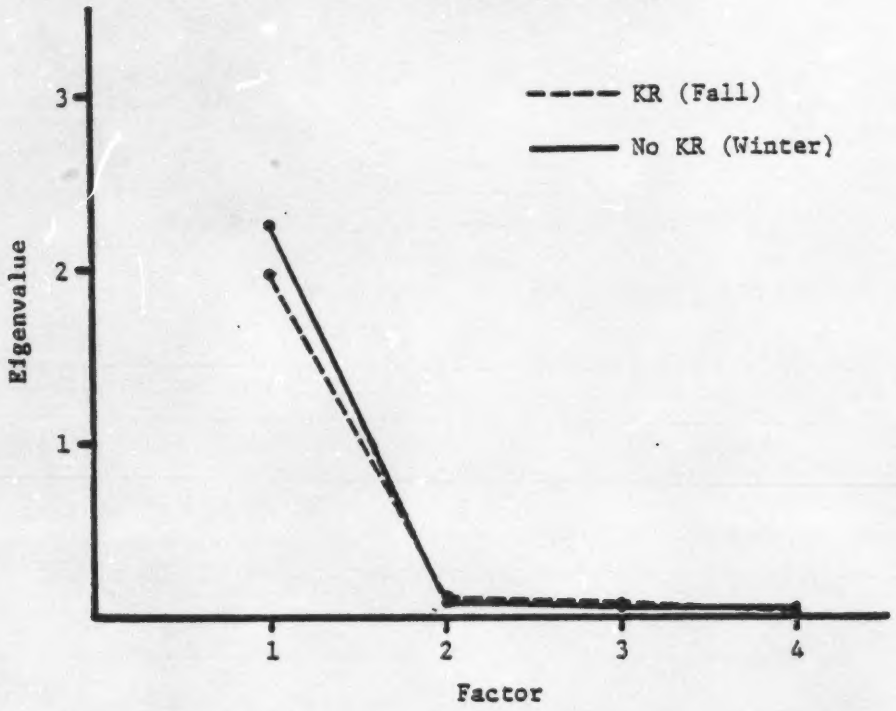
Principal axes factor analyses of the subtest scores were performed separately for each adaptive midquarter exam. Table 4 shows the results when all four factors were extracted from the data.

Table 4
Eigenvalues and Percent of Variance Accounted for when Four Factors were Extracted from Subtest Intercorrelations for Exams 1 and 2, for KR (Fall) and No-KR (Winter) Groups

Exam and Factor	KR (Fall)			No-KR (Winter)		
	Eigenvalue	% Common Variance	% Total Variance	Eigenvalue	% Common Variance	% Total Variance
Exam 1						
Factor 1	1.98	91.6	50.0	2.28	93.0	57.0
Factor 2	.10	4.7	2.6	.08	3.2	2.0
Factor 3	.07	3.4	1.8	.05	2.0	1.2
Factor 4	.01	.3	.2	.05	1.8	1.1
Total	—	100.0	54.6	—	100.0	61.3
Exam 2						
Factor 1	2.50	94.7	62.6	2.34	93.3	58.6
Factor 2	.07	2.8	1.9	.09	3.5	2.2
Factor 3	.05	2.0	1.3	.04	1.7	1.1
Factor 4	.01	.6	.4	.04	1.6	1.0
Total	—	100.0	66.2	—	100.0	62.9

Exam 1. Table 4 shows that the four principal factors extracted from the data of the KR group accounted for 54.6% of the total variance of the subtest scores. The first of these factors accounted for nearly 50% of the total variance, or 91.6% of the common variance. The remaining three factors accounted for relatively small amounts of the remaining variance, as shown in Figure 2, which is a plot of the eigenvalues versus the factors extracted and illustrates the relative unimportance of the last three factors extracted.

Figure 2
Eigenvalues for the Four Factors of Exam 1 for KR (Fall)
and No-KR (Winter) Groups



The results observed for Exam 1 from the no-KR group were very similar. As Table 4 and Figure 2 show, one factor accounted for the majority (53.0%) of the common variance, with the last three factors accounting for very little variance. The four factors extracted from this exam accounted for a greater amount of the total variance (61.3%) than did the factors extracted from the KR group data; the first factor extracted accounted for 57.0% of the total variance of the subtest scores.

It appears, then, that the first factor extracted from these two exams was stronger in the case where no KR was administered to the students, and it accounted for a greater proportion of the common variance. Equivalently, when immediate KR was administered (fall), the second and third factors were stronger than when KR was absent (winter).

This same phenomenon can be observed in the patterns of factor loadings and communalities, as given in Table 5. The communality estimates and the factor loadings for the first factor were higher for the no-KR group than they were for the KR group. For the no-KR group the first factor loadings ranged from .73 to .79; for the KR group they ranged from .67 to .73.

In both cases, the factor loadings for the first factor were smallest for Subtest 1 and largest for Subtest 4. This was to be expected because of the way items were selected for administration. That is, although each subtest was scored independently of the others (e.g., scoring for the first item of each subtest began by assuming a

Table 5
Factor Loadings and Communality Estimates for Subtests of Exam 1 when Four Factors were Extracted for KR (Fall) and No-KR (Winter) Groups

Group and Subtest	Factor 1	2	3	4	Communality Estimate
KR (Fall)					
Subtest 1	.67	.18	.13	.04	.50
Subtest 2	.72	-.17	-.12	.04	.56
Subtest 3	.71	-.15	.15	-.04	.54
Subtest 4	.73	.15	-.14	-.04	.58
No-KR (Winter)					
Subtest 1	.73	.14	.15	.02	.58
Subtest 2	.75	.14	-.15	.00	.60
Subtest 3	.75	-.15	.00	.14	.60
Subtest 4	.79	-.12	.01	-.16	.67

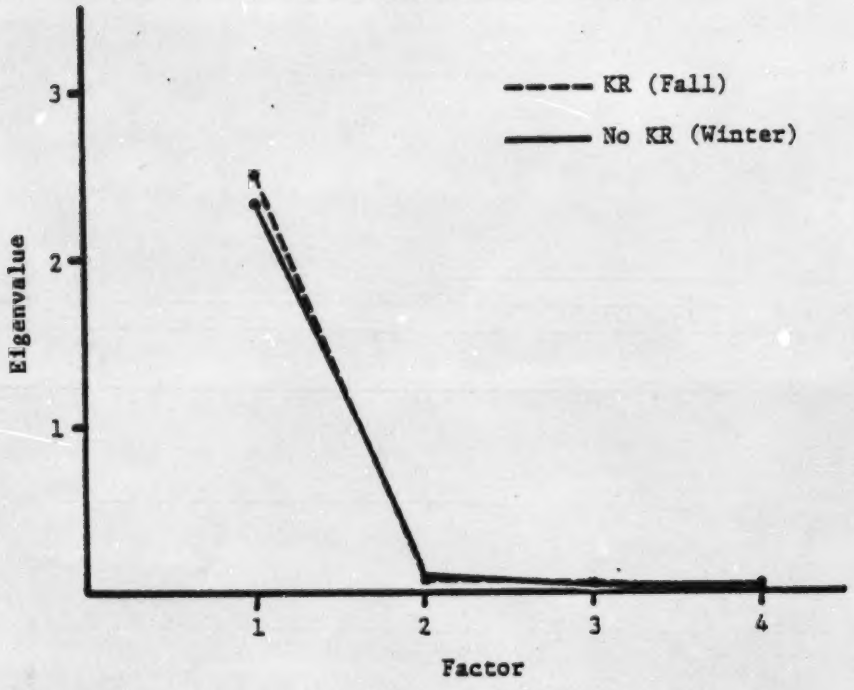
normal prior distribution of achievement with a mean of 0 and a standard deviation of 1.0), because the test was adaptive, administration of each item was dependent upon the student's response to the previous item. As the test progressed, the items administered were closer in difficulty to the student's true achievement level; hence, there should be less error involved than at the begining of the test. This should translate into higher loadings on the first biology factor. The communality estimates for each subtest also showed the same trends.

Exam 2. Virtually the same pattern of factor loadings and eigenvalues was observed for Exam 2, with a few important exceptions. As shown in Table 4, the four principal axis factors extracted from the KR group data accounted for 66.2% of the total variance of the subtest scores, with the first of these factors accounting for 62.6% of the total variance (and 94.7% of the common variance). As before, the three remaining factors accounted for little of the remaining variance.

The results observed for the no-KR group were very similar to those obtained for the KR group. As Table 4 shows, one factor accounted for the majority (93.3%) of the common variance, with the last three factors accounting for very little. The four factors extracted from this exam accounted for a lesser amount of the total variance than did the factors extracted from the KR group data (62.8% versus 66.2%). For the no-KR group the first factor accounted for 58.55% of the total variance of the subtest scores.

Figure 3 is a plot of the eigenvalues versus the factors extracted, for Exam 2. In contrast to the results observed for Exam 1 (Figure 2), the first factor extracted from Exam 2 was slightly stronger in the case when KR was administered (fall), and it accounted for a greater proportion of the total variance. On the other hand, when KR was absent (winter), the second factor was slightly stronger than when KR was administered.

Figure 3
Eigenvalues for the Four Factors of Exam 2 for KR (Fall) and No-KR (Winter) Groups



This same phenomenon was apparent in the patterns of factor loadings and communality estimates shown in Table 6. The factor loadings for the first factor from the KR (fall) group were somewhat higher than the corresponding factor loadings for the no-KR (winter) group. In the KR group they ranged from .72 to .87, as compared with

Table 6
Factor Loadings and Communality Estimates for Subtests of Exam 2 when Four Factors were Extracted for KR (Fall) and No-KR (Winter) Groups

Group and Subtest	Factor 1	2	3	4	Communality Estimate
KR (Fall)					
Subtest 1	.75	-.17	-.07	.06	.59
Subtest 2	.72	.18	.05	.07	.55
Subtest 3	.82	-.08	.16	-.05	.71
Subtest 4	.87	.07	-.13	-.07	.79
No-KR (Winter)					
Subtest 1	.74	.15	.14	.03	.59
Subtest 2	.77	.14	-.15	-.02	.63
Subtest 3	.76	-.16	-.02	.14	.61
Subtest 4	.80	-.13	.03	-.14	.67

.74 and .80 for the no-KR group. These data show greater variance in the first factor loadings for the KR group as compared to the no-KR group. This same phenomenon was observed in the communality estimates for the four subtests from each quarter. That is, the communalities in the KR data ranged from .55 to .79; those in the no-KR data varied much less, from .59 to .67. As before, however, the highest factor loadings and the highest communality estimates were observed for the last of the four subtests.

Conclusions

Subtest Scores

No consistent differences in achievement level estimates were observed between the KR and no-KR groups. There were slightly lower mean achievement level scores for students who were administered KR in Exam 2. Because this difference was fairly constant across subtests, there was no evidence in these data of the occurrence of cumulative KR effects. Rather, the only slight (but nonsignificant) effect appeared to be a general lowering of scores on the entire test. Because the item pools differed for the two quarters, however, it is not known to what extent this small observed group difference was artifactual.

Subtest Intercorrelations and Factor Analyses

The patterns of inter-subtest correlations reflected (not surprisingly) the results from the factor analyses. For Exam 1 the no-KR group correlations were more homogeneous, and higher, than were the correlations in the KR group. The factor loadings for the first factor from the no-KR group were higher, the communality estimates for each subtest were higher, and this first factor accounted for a greater proportion of both the total and common variance when compared to the KR group.

For Exam 2 it was the correlations in the KR group that were more homogeneous; there were no mean differences in the sizes of the correlations between the two groups. Similarly, a stronger first factor was observed in Exam 1 for the KR group.

If administration of immediate KR after each item provided information that aided a student in answering subsequent items, then the assumption of local independence among the items should have been violated. Such a violation should result in differing factor patterns for the two quarters. However, such a result did not occur. Although there were differences in the patterns of factor loadings and eigenvalues between the two groups, these differences were small and not consistently in the same direction. That is, inspection of the Exam 1 data alone might lead to the conclusion that the administration of KR violates the assumption of unidimensionality to a certain extent. Inspection of the Exam 2 data, however, revealed just the opposite: The first factor was, in fact, stronger when KR was administered than when it was not.

On the basis of this study alone, it may be concluded that administration of immediate KR does not violate the latent trait as-

sumption of unidimensionality and local independence across achievement test items; nor does it seem to raise or to lower test scores to any significant degree. The comparisons made here between KR and no KR, however, may be confounded to some unknown degree because the item pools differed slightly across quarters and because the data were collected from different academic quarters. If either of these variables were salient to the question at hand, the results reported here might be misleading. Study II attempted to replicate these findings using a different experimental design.

STUDY II

Method

Subjects

Data for this study were obtained from students enrolled in the same introductory biology course at the University of Minnesota during winter quarter 1978. As in Study I, volunteers were recruited to take computerized adaptive versions of their two classroom midquarter exams. Students in Study II also received extra credit points for their participation. The first computerized adaptive midquarter exam (Exam 1) was administered to 252 students; the second computerized adaptive midquarter exam (Exam 2) was administered to 305 students.

Procedure

For both the first and second adaptive test administrations, the students were given their adaptive biology test followed by a 45-item vocabulary test. All tests were computer administered, following the same procedures used in Study I.

For each of the tests, students were sequentially assigned to one of two testing conditions: KR or no KR. That is, approximately half the students received KR immediately following each response; the other half did not. There were no time limits imposed on any of the tests. At the completion of testing, students received a printed report, listing items answered incorrectly along with the correct answers.

Adaptive Test

The item pools used for this study were the item pools used for the winter quarter in Study I. The testing strategy (i.e., stradaptive) and determination of entry stratum were exactly as described in Study I. The only difference between the procedures for these two studies was that the maximum number of items was set at 40 (instead of 50) for this study. Following the same procedures used in Study I, the adaptive tests were divided into four subtests of equal length, these subtests were scored, and principal axes factor analyses were performed on the subtest scores.

Results

Subtest Scores

The means and standard deviations of the subtest achievement level estimates ($\hat{\theta}$) for each of the computerized adaptive midquarter exams (Exam 1 and Exam 2) are given in Table 7; the mean achievement level estimates are plotted, separately for each exam, in Figure 4.

Table 7
Sample Size (N), Means, and Standard Deviations (S.D.) of Subtest Achievement Level Estimates for Exams 1 and 2, for KR and No-KR Groups

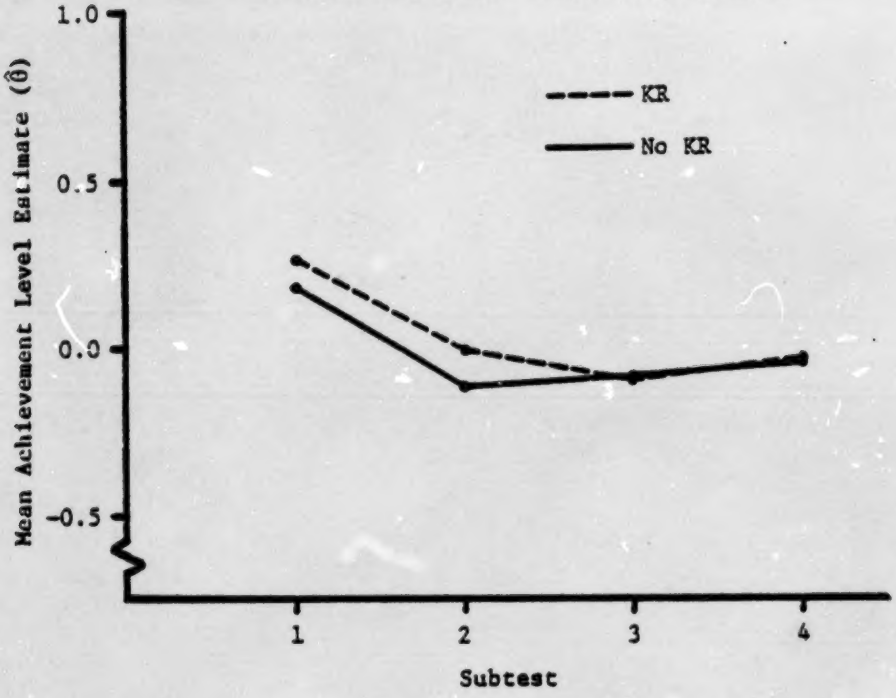
Exam and Subtest	KR N	KR Mean	KR S.D.	No-KR N	No-KR Mean	No-KR S.D.
Exam 1	138			114		
Subtest 1		.26	1.07		.18	1.03
Subtest 2		-.01	.82		-.12	.66
Subtest 3		-.10	.80		-.08	.77
Subtest 4		-.03	.83		-.05	.82
Exam 2	151			154		
Subtest 1		-.27	.74		-.36	.70
Subtest 2		-.19	.64		-.29	.74
Subtest 3		-.23	.65		-.30	.63
Subtest 4		-.25	.60		-.32	.67

Inspection of this table and figure shows that there were some consistent differences between the KR and no-KR groups in the achievement level estimates obtained for Exam 1 and Exam 2. For Exam 1 (Figure 4a), students who received KR had slightly (but nonsignificantly, at the .05 level) higher estimated achievement levels when compared to students who had not received KR for the first two subtests; for the last two subtests, the performance of the two groups was indistinguishable. For Exam 2, students who received KR also had higher estimated achievement levels than those students who did not. Although this discrepancy was small (approximately .1 unit on the θ metric) and nonsignificant ($p=.05$), it was consistent (not cumulative) across all four subtests.

Subtest Intercorrelations

The subtest intercorrelation matrices for the winter 1978 midquarter exams are given in Tables 8 and 9. Table 8 shows these correlations for Exam 1 for the students who received KR (above diagonal) and for those who did not (below diagonal). The correlations within each group were fairly heterogeneous, ranging from .47 to .68 for the KR group and from .51 to .70 for the no-KR group. For this exam, the inter-subtest correlations for the no-KR group were slightly, though uniformly, higher than the inter-subtest correlations for the KR group.

Figure 4
Mean Achievement Level Estimates ($\hat{\theta}$) for KR and No-KR (a) Exam 1 Groups, by Subtest



(b) Exam 2

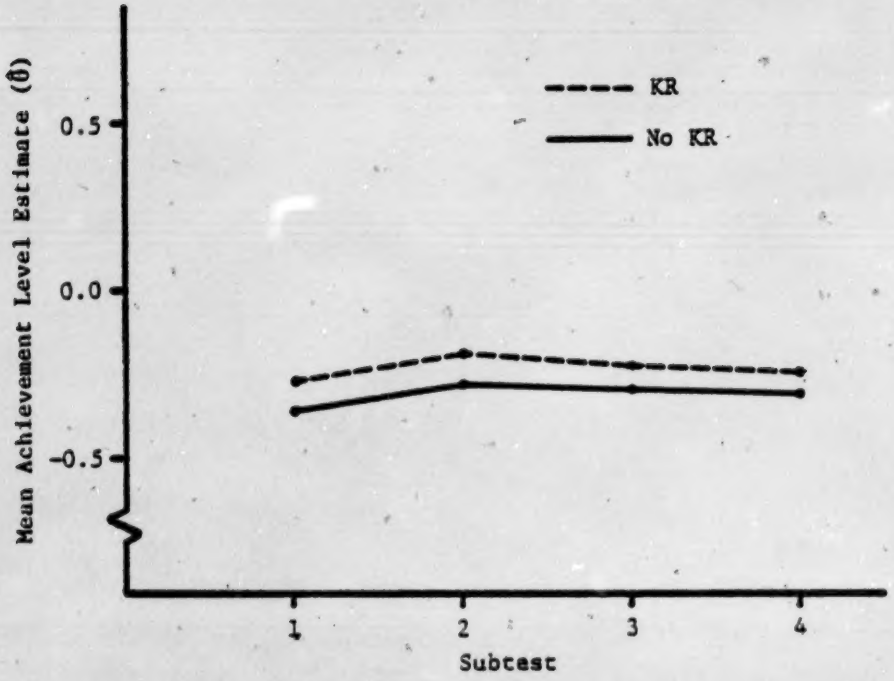


Table 8
Exam 1 Subtest Intercorrelations for KR Group (Above Diagonal) and No-KR Group (Below Diagonal)

	Subtest			
Subtest	1	2	3	4
1		.47	.56	.63
2	.51		.48	.46
3	.56	.56		.68
4	.70	.59	.62	

Table 9 presents the inter-subtest correlations from the two groups of students for Exam 2. These correlations were more homogeneous and generally lower than the correlations for Exam 1. For

Table 9
Exam 2 Subtest Intercorrelations for KR Group (Above Diagonal) and No-KR Group (Below Diagonal)

	Subtest			
Subtest	1	2	3	4
1		.51	.48	.55
2	.46		.49	.46
3	.40	.43		.54
4	.57	.43	.50	

this exam they ranged from .46 to .55 for the KR group and from .40 to .57 for the no-KR group. The no-KR group correlations varied more than did the KR group correlations. Again, as in Study I, the generally low level of subtest intercorrelations resulted from the average subtest length of about eight items.

Factor Analyses

Principal axes factor analyses of the subtest scores were performed separately for each exam. The results, when all four factors were extracted from the data, are given in Table 10.

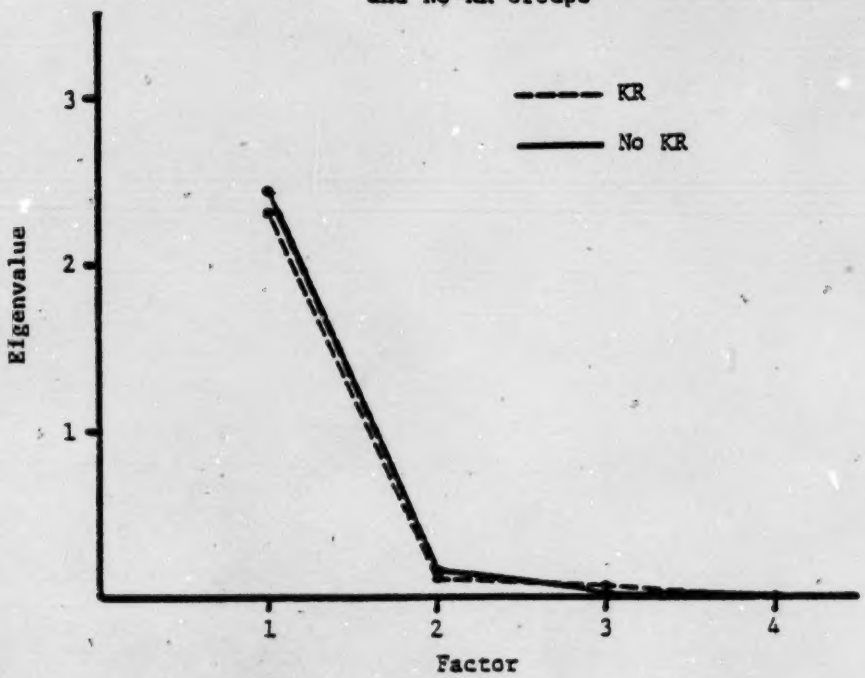
Exam 1. Table 10 shows that the four principal factors extracted from the exams in the KR group accounted for 62.1% of the total variance of the subtest scores. The first of these factors accounted for over 57% of the total variance, or 92.3% of the common variance. The remaining three factors accounted for relatively small amounts of the remaining variance, as shown in Figure 5, which is a plot of the eigenvalues versus the factors extracted.

Table 10
Eigenvalues and Percent of Variance Accounted for when Four Factors were Extracted from Subtest Intercorrelations for Exams 1 and 2, for KR and No-KR Groups

Exam and Factor	KR			No-KR		
	Eigen-value	% Common Variance	% Total Variance	Eigen-value	% Common Variance	% Total Variance
Exam 1						
Factor 1	2.29	92.3	57.3	2.43	94.0	60.8
Factor 2	.11	4.6	2.8	.13	5.1	3.3
Factor 3	.08	3.2	2.0	.02	.9	.6
Factor 4	.00	.0	.0	.00	.1	.0
Total	—	100.0	62.1	--	100.0	64.7
Exam 2						
Factor 1	2.09	91.0	52.1	1.96	87.7	49.1
Factor 2	.11	4.6	2.7	.15	6.9	3.9
Factor 3	.10	4.4	2.5	.12	5.4	3.0
Factor 4	.00	.1	.0	.00	.1	.0
Total	--	100.0	57.3	--	100.0	56.0

The results observed for the no-KR group were very similar. As Table 10 and Figure 5 show, one factor accounted for nearly 94% of the total variance, with the last three factors accounting for very little variance. The four factors extracted from this exam in the

Figure 5
Eigenvalues for the Four Factors of Exam 1 for KR and No-KR Groups



no-KR group accounted for a slightly greater amount of the total variance (64.7%) than did the factors extracted for the KR group (62.1%); the first factor extracted in the no-KR group accounted for 60.84% of the total variance of the subtest scores.

It appears, then, that the first factor extracted from Exam 1 in these two groups was slightly stronger in the case where no KR was administered to the students, and it accounted for a greater proportion of the total variance. On the other hand, when immediate KR was administered, the third factor was stronger than when KR was absent.

This same phenomenon is illustrated in the patterns of factor loadings and communalities, as given in Table 11. The communality estimates and the factor loadings for the first factor were all higher for the no-KR group than they were for the KR group. For the

Table 11
Factor Loadings and Communality Estimates for Subtests when Four Factors were Extracted from Exam 1 for KR and No-KR Groups

Group and Subtest	Factor 1	2	3	4	Communality Estimate
KR					
Subtest 1	.76	.17	-.14	-.02	.63
Subtest 2	.60	.19	.14	.02	.42
Subtest 3	.80	-.14	.15	-.02	.69
Subtest 4	.84	-.16	-.12	.02	.75
No-KR					
Subtest 1	.78	-.19	.07	.02	.65
Subtest 2	.71	.20	-.07	.02	.55
Subtest 3	.75	.19	.08	-.02	.61
Subtest 4	.87	-.15	-.07	-.02	.78

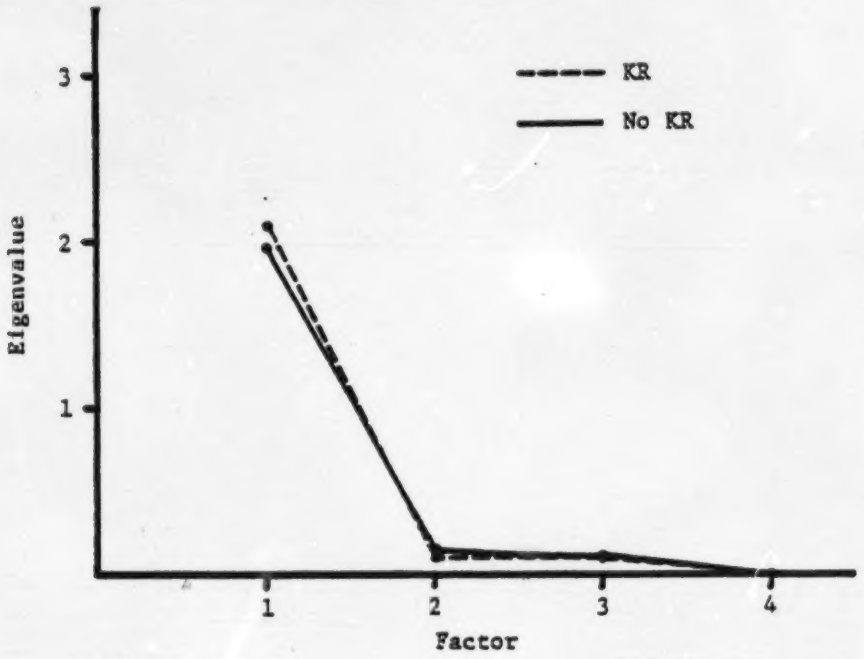
no-KR group, the first factor loadings ranged from .71 to .87; for the KR group, the range was from .60 to .84.

In both groups the factor loadings for the first factor were smallest for the second subtest and reached their largest value for the fourth subtest. Although the drop in the size of the factor loadings for Subtest 2 was surprising, the higher factor loading for Subtest 4 was expected because of the way items were selected for administration. The communality estimates for each subtest showed the same trends.

Exam 2. A similar pattern of factor loadings and eigenvalues was observed for Exam 2, with a few important exceptions. As shown in Table 10, the four principal axes factors extracted from the KR group accounted for 57.3% of the total variance of the subtest scores, with the first of these factors accounting for 52.1% of the total variance (and 91.0% of the common variance). As before, the

three remaining factors accounted for little of the remaining variance. Figure 6 is a plot of the eigenvalues versus the factors extracted for Exam 2 and illustrates once again the relative unimportance of the last three factors extracted.

Figure 6
Eigenvalues for the Four Factors of Exam 2 for KR and No-KR Groups



The results observed for the no-KR group were very similar to those for the KR group. As Table 10 and Figure 6 show, one factor accounted for the majority (67.7%) of the common variance, with the last three factors accounting for relatively little variance. The four factors extracted from this exam accounted for a slightly lesser amount of the total variance than did the factors extracted from the exams for the KR group (56.0% versus 57.3%); the first factor in the no-KR group accounted for 49.1% of the total variance of the subtest scores.

In contrast to what was observed for Exam 1, the first factor extracted from these two exams was a bit stronger in the case when KR was administered, and it accounted for a greater proportion of the total variance. Equivalently, when KR was absent, the second and third factors were stronger than when it was administered.

The patterns of factor loadings and communality estimates for the KR and no-KR groups from Exam 2 are given in Table 12. The factor loadings for the first factor extracted from the KR group were less variable than from the no-KR group. For the KR group they ranged from .68 to .75; for the no-KR group the factor loadings varied between .64 and .78. This same result was observed in the commu-

nality estimates for the four subtests from each group. That is, although the communalities for the KR group ranged from .52 to .61, those for the no-KR group varied much more, from .48 to .66. As before, however, the highest factor loadings and the highest communality estimates were observed for the last of the four subtests.

Table 12
Factor Loadings and Communality Estimates for Subtests when Four Factors were Extracted from Exam 2 for KR and No-KR Groups

Group and Subtest	Factor 1	2	3	4	Communality Estimate
KR					
Subtest 1	.74	.12	-.16	-.02	.60
Subtest 2	.68	.20	.12	.02	.52
Subtest 3	.72	-.12	.20	-.02	.56
Subtest 4	.75	-.19	-.11	.02	.61
No-KR					
Subtest 1	.73	-.21	.15	.02	.60
Subtest 2	.64	.18	.20	-.02	.48
Subtest 3	.65	.23	-.15	.02	.50
Subtest 4	.78	-.15	-.19	-.02	.66

Conclusions

Subtest Scores

In this study some consistent (but, again, nonsignificant) differences in achievement level estimates were observed when the KR group was compared to the no-KR group. Whenever a difference between the groups existed, it favored the group of students who received immediate KR. Such differences were observed for the first two subtests of Exam 1 and for all four subtests of Exam 2. For Exam 2 this difference was fairly constant across subtests, with no indication that the effects of KR were cumulative over the contiguous item sets within a test. These data lead to the tentative conclusion that immediate KR in an achievement test results in slightly, though uniformly, higher subtest scores.

Subtest Intercorrelations and Factor Analyses

For Exam 1 the inter-subtest correlations were fairly heterogeneous within each group, with higher correlations observed for the no-KR group than for the KR group. The factor loadings for the first factor from the no-KR group were higher, the communality estimates for the subtests were higher, and this first factor accounted for a greater proportion of both the total and common variance when compared to the KR group.

For Exam 2 the inter-subtest correlations were more homogeneous for the KR group than for the no-KR group, although there were no

mean differences in the sizes of the correlations between the groups. Similarly, a stronger first factor was observed for the KR group.

Although there were differences in the patterns of factor loadings and eigenvalues between the two groups, these differences were small and not consistently in the same direction. Hence, there was no data in this study to support the conclusion that the provision of immediate KR violated the assumption of local independence among the items.

DISCUSSION AND CONCLUSIONS

The primary question under investigation in both studies was, "If students are told the correct answers to items administered early in an achievement test, will that information affect their responses to subsequent test items?" The answer to this question has, essentially, two distinct components.

Subtest Scores

First, it may be asked whether the provision of immediate KR affects a student's test score to any significant degree. If the correct answers to some items contained clues or information regarding the correct answers to later items, then it is reasonable to expect that students who received such informative KR would, on the average, have higher test scores than students who did not. Further, such effects may well be cumulative in nature; the differences between groups of students who received informative KR and those who did not may be larger for items administered toward the end of the test.

The results from these two studies with regard to this question are equivocal at best. For Exam 1 the achievement level estimates for the two groups were essentially the same. In Study 2 the KR group scored higher, on the average, than did the no-KR group, although this difference was observed early in the test (for the first two subtests) only. For Exam 2 consistent differences between the groups were observed in both studies. For Study I the advantage was for the no-KR group; for Study II the advantage was for the KR group. The differences observed in both studies were constant across subtests; that is, instead of having an effect that was cumulative in nature, administration of KR resulted in a constant general elevation or depression of scores across subtests. Coupled with the fact that all observed differences were slight (e.g., .1 unit on the θ metric), this leads to the conclusion that administration of informative KR has no meaningful effect on achievement test scores.

Factor Analyses

The second, perhaps more important, question motivating these two studies was whether or not the provision of immediate KR affected the dimensionality of achievement tests. If the correct answers to some items contained clues or information regarding the correct answers to later items, then it is reasonable to expect that the items are (by definition) no longer locally independent. This should re-

sult in a difference in the factor structures of the exams between the two groups, for example, a weaker first factor and/or stronger second factor for the KR groups.

Again, the results from these two studies were inconsistent when comparisons between groups were made. The data from Exam 1 (for both studies) were as expected. That is, the first factor from the exam, when administered to the no-KR groups was stronger than the first factor from the same exam administered to the KR group, in the sense of accounting for a larger proportion of the total and common variance. Also, there was more common variance for the no-KR groups than for the KR groups. These results would be expected if the administration of KR violated the assumption of unidimensionality or local independence among the items. Yet the observed differences between the KR and no-KR groups in proportion of variance accounted for by the common factors were slight in all cases.

The results from the Exam 2 data, however, pointed to the opposite conclusion. For these data, in both studies, the stronger first factor was observed for the KR groups; and the full solutions accounted for a smaller proportion of the total variance for the no-KR groups. Again, differences between the groups, though consistent, were slight.

On the basis of these data, then, it may be concluded that administration of immediate KR does not violate the latent trait model assumption of unidimensionality or local independence among achievement test items.

Further Research

Perhaps the most fruitful direction for further research into this issue is replication of these studies with the analyses done at the item level. Because the data for these studies were responses to an adaptive test, the unit of analysis here was subtest scores. It is possible that these scores are simply too global in nature to be sensitive to subtle changes in test dimensionality and score level.

Also, analyses on the item level may yield information concerning which item responses (if any) were most influenced by information contained in previous items. The degree to which the dimensionality of a test is affected by the administration of KR may depend very heavily on the type of items included in the test and on the number of items directly influenced by the information contained in previously administered items.

This issue should also be investigated in the context of the particular subject matter and item type to be used. However, these studies indicate that administration of informative KR does not alter the dimensionality of an achievement test nor affect test score in any systematic or meaningful way. This implies that latent trait models may be appropriate for use in achievement tests even when informative KR is administered immediately after each item response.

REFERENCES

Bejar, I. I., Weiss, D. J., & Kingsbury, G. G. Calibration of an item pool for the adaptive measurement of achievement (Research Report 77-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1977. (NTIS No. AD A044828)

Betz, N. E. Prospects: New types of information and psychological implications. In D. J. Weiss (Ed.), Computerized adaptive trait measurement: Problems and prospects (Research Report 75-5). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, November 1975. (NTIS No. AD A018675)

Betz, N. E., & Weiss, D. J. Effects of immediate knowledge of results and adaptive testing on ability test performance (Research Report 76-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, June 1976. (NTIS No. AD A028147) (a)

Betz, N. E., & Weiss, D. J. Psychological effects of immediate knowledge of results and adaptive ability testing (Research Report 76-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, June 1976. (NTIS No. AD A027170) (b)

Birnbaum, A. Some latent trait models and their use in inferring an examinee's ability. In F. M. Lord & M. R. Novick, Statistical theories of mental test scores. Reading, MA: Addison-Wesley, 1968.

Cronbach, L. J. Essentials of psychological testing (3rd ed.). New York: Harper & Row, 1970.

De Witt, C. J., & Weiss, D. J. A computer software system for adaptive ability measurement (Research Report 74-1). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, January 1974. (NTIS No. AD 773961)

Owen, R. J. A Bayesian sequential procedure for quantal response in the context of adaptive mental testing. Journal of the American Statistical Association, 1975, 70, 351-356.

Samuda, R. J. Psychological testing of American minorities: Issues and consequences. New York: Dodd, Mead, & Co., 1975.

Terman, L. M. The measurement of intelligence. New York: Houghton-Mifflin, 1916.

Vale, C. D., & Weiss, D. J. A study of computer-administered stradaptive ability testing (Research Report 75-4). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, October 1975. (NTIS No. AD A018758) (a)

Vale, C. D., & Weiss, D. J. A simulation study of stradaptive ability testing (Research Report 75-6). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, Decembe: 1975. (NTIS No. AD A020961) (b)

Weiss, D. J. The stratified adaptive computerized ability test (Research Report 73-3). Minneapolis: University of Minnesota, Department of Psychology, Psychometric Methods Program, September 1973. (NTIS No. AD 768376)

APPENDIX: SUPPLEMENTARY TABLES

Table A
Item Number, Discrimination (*a*), Difficulty (*b*), and Guessing (*c*) Parameters for Items in the Midquarter 1 Stradaptive Item Pool

Item	*a*	*b*	*c*
Stratum 9 (15 items)			
3209	2.50	2.29	.29
3417	2.50	3.00	.35
3033	1.54	2.44	.35
3440*	1.52	2.00	.30
3231	2.50	2.39	.35
3406	1.31	2.48	.35
3045	1.02	2.48	.27
3242	.94	2.40	.35
3407	1.02	2.41	.29
3263*	.99	2.29	.35
3241	.91	2.09	.17
3414	.88	2.29	.32
3402	.83	2.44	.35
3247	.82	2.42	.35
3228	.67	2.49	.31
Mean (F)	1.34	2.43	.32
Mean*(W)	1.33	2.39	.32
Stratum 8 (20 items)			
3409	2.50	1.28	.00
3234	2.50	1.73	.00
3018	.89	1.25	.35
3204	1.14	1.66	.35
3422	1.47	1.50	.35
3411	1.36	1.23	.35
3250	.91	1.94	.29
3206	.74	1.51	.21
3410	1.30	1.34	.31
3429	1.25	1.24	.28
3419	1.23	1.48	.25
3421	1.17	1.15	.35
3436*	1.12	1.59	.35
3271*	.95	1.32	.30
3061*	.95	1.57	.30
3427	.92	1.51	.26
3449*	.91	1.26	.14
3063*	.91	1.51	.35
3074*	.84	1.79	.35
3420	.68	1.62	.35
Mean (F)	1.29	1.46	.26
Mean*(W)	1.19	1.47	.27
Stratum 7 (20 items)			
3408	2.50	1.05	.31
3437*	1.95	.66	.28
3258	1.24	.81	.35
3432	1.72	.67	.35
3048	1.35	.66	.33
3413	1.40	.76	.35
3448*	1.40	.73	.30
3439*	1.36	.64	.32
3219	1.23	.62	.21
3072*	1.02	.65	.32
3277*	1.00	1.04	.35
3035	.90	.68	.28
3433	1.35	.86	.30
3447*	1.18	.93	.32
3064*	.94	.86	.24
3230	.90	.87	.35
3444*	.88	.78	.35
3012	.75	.80	.35
3260	.71	.84	.28
3056*	.71	.89	.26
Mean (F)	1.28	.78	.31
Mean*(W)	1.22	.79	.31
Stratum 6 (19 items)			
3047	1.66	.44	.29
3079*	1.61	.27	.35
3213	.93	.52	.35
3041	1.51	.23	.35
3062*	1.47	.43	.30
3405	1.40	.55	.32
3445*	1.19	.44	.34
3218	.82	.58	.12
3019	1.31	.29	.29
3207	.70	.46	.28
3431	.70	.28	.34
3000	1.24	.52	.35
3046	1.18	.24	.22
3042	1.15	.37	.27
3050	1.13	.35	.18
3066*	1.05	.53	.31
3034	1.01	.37	.28
3262*	.81	.47	.35
3438*	.70	.21	.27
Mean (F)	1.13	.40	.28
Mean*(W)	1.14	.40	.29
Stratum 5 (15 items)			
3282*	2.06	-.02	.35
3220	1.79	-.03	.26
3005	1.43	.11	.35
3425	1.36	.17	.23
3053	1.12	.12	.00
3214	1.12	.03	.23
3412	1.12	.19	.35
3051	1.29	.21	.28
3279*	.99	.01	.28
3403	.99	.18	.19
3069*	.88	-.01	.35
3211	.88	.01	.13
3002	.82	.13	.14
3426	.68	.07	.22
3423	.66	.16	.27
Mean (F)	1.11	.11	.22
Mean*(W)	1.15	.09	.24
Stratum 4 (13 items)			
3256	2.31	-.33	.26
3430	1.15	-.30	.29
3031	1.47	-.33	.35
3254	3.38	-.17	.22
3237	1.54	-.37	.18
3404	.65	-.29	.35
3244	1.35	-.44	.23
3058*	1.05	-.43	.35
3240	.98	-.28	.15
3268*	.97	-.28	.18
3208	.76	-.16	.12
3006	.77	-.37	.33
3259	.69	-.41	.20
Mean (F)	1.27	-.31	.25
Mean*(W)	1.23	-.32	.25
Stratum 3 (19 items)			
3021	1.96	-.49	.21
3217	1.06	-.48	.14
3052	1.71	-.93	.00
3055*	1.71	-.65	.24
3215	1.59	-.82	.23
3011	1.32	-.86	.20
3435*	.83	-.61	.35
3216	1.27	-.62	.18
3054*	1.29	-.93	.31
3221	1.25	-.52	.17
3049	1.15	-.71	.18
3255	1.14	-.72	.26
3067*	1.07	-.76	.21
3246	1.10	-.72	.28
3022	1.01	-.48	.30
3272*	1.06	-.81	.35
3017	.99	-.58	.16
3076*	.94	-.73	.21
3224	.80	-.50	.37
Mean (F)	1.26	-.65	.20
Mean*(W)	1.22	-.68	.22
Stratum 2 (20 items)			
3023	2.40	-1.15	.35
3202	1.81	-.99	.21
3415	.85	-.96	.35
3245	1.34	-.96	.21
3236	1.26	-1.20	.33
3020	1.23	-1.28	.17
3028	1.12	-1.26	.35
3226	1.09	-.98	.20
3210	1.04	-1.22	.35
3239	1.04	-1.13	.21
3013	1.00	-.97	.35
3267*	1.02	-1.22	.23
3257	.98	-1.02	.25
3070*	.95	-1.28	.22
3036	.92	-1.18	.16
3014	.86	-1.24	.14
3060*	.86	-1.3?	.29
3274*	.85	-1.05	.26
3238	.82	-1.06	.21
3032	.77	-1.06	.27
Mean (F)	1.16	-1.10	.26
Mean*(W)	1.11	-1.13	.26
Stratum 1 (17 items)			
3077*	2.50	-1.39	.20
3027	1.67	-1.38	.35
3443*	1.07	-1.64	.35
3249	.91	-1.69	.17
3428	.90	-1.56	.35
3073*	1.43	-1.57	.31
3205	1.25	-1.53	.19
3078*	1.24	-1.65	.35
3057*	1.20	-1.35	.26
3065*	1.17	-1.66	.35
3235	1.15	-1.40	.28
3029	1.13	-1.50	.28
3201	1.07	-1.34	.23
3008	.96	-1.75	.18
3252	.79	-1.77	.35
3003	.96	-1.76	.34
3044	.87	-1.42	.15
Mean (F)	1.06	-1.55	.26
Mean*(W)	1.19	-1.55	.28

Note. Items with asterisks are those which were added to the pool Winter quarter. All other items were in the pool both Fall and Winter quarters.

Table 8
Item Number, Discrimination (a), Difficulty (b), and Guessing (c) Parameters for Items in the Midquarter 2 Stradaptive Item Pool

Item	a	b	c	Item	a	b	c	Item	a	b	c
Stratum 9 (18 items)				Stratum 6 (20 items)				Stratum 3 (17 items)			
3831	2.50	1.96	.06	3707*	1.75	.55	.31	3634	1.79	-.58	.30
3690	2.50	2.36	.24	3746*	1.59	.43	.30	3739*	1.68	-.61	.35
3833*	2.50	2.85	.35	3806	1.57	.48	.35	3809	1.27	-.61	.35
3904	2.45	1.48	.28	3925*	1.14	.48	.35	3924*	1.13	-.79	.18
3805	2.30	2.38	.35	3638	1.24	.32	.35	3672	1.57	-.80	.15
3698	2.11	2.82	.35	3905	.98	.35	.20	3737*	1.41	-.66	.34
3901	1.35	2.62	.35	3738*	1.34	.40	.35	3915	1.08	-.61	.16
3835*	1.21	2.28	.35	3605	1.22	.57	.34	3640	1.43	-.69	.35
3620	2.04	2.97	.35	3815	.95	.58	.35	3906	.87	-.66	.14
3697	1.56	3.00	.35	3611	1.22	.39	.32	3812	.82	-.63	.13
3810	.92	2.20	.27	3675	1.21	.40	.28	3682	1.33	-.72	.34
3664	1.11	1.60	.35	3820	.92	.38	.12	3637	1.29	-.73	.28
3625	.98	1.66	.35	3665	1.19	.54	.22	3636	1.24	-.63	.27
3622	.95	2.53	.35	3709*	1.19	.30	.35	3641	1.20	-.65	.22
3841*	.87	2.13	.35	3724*	1.14	.37	.30	3711*	1.05	-.56	.35
3651	.95	2.30	.35	3819	.76	.53	.35	3608	1.04	-.78	.16
3728*	.91	2.55	.35	3918*	.66	.35	.23	3705*	.87	-.58	.14
3712*	.75	1.64	.30	3614	.79	.46	.35				
				3923*	.63	.38	.31	Mean (F)	1.24	-.67	.24
Mean (F)	1.70	2.31	.31	3626	.65	.32	.25	Mean*(W)	1.25	-.66	.25
Mean*(W)	1.58	2.30	.32								
				Mean (F)	1.06	.46	.29	Stratum 2 (20 items)			
Stratum 8 (18 items)				Mean*(W)	1.11	.44	.30	3733*	1.63	-.94	.35
3613	1.69	1.17	.29	Stratum 5 (15 items)				3648	1.59	-.96	.33
3916	1.39	1.14	.35	3742*	1.89	.27	.35	3807	1.32	-1.10	.17
3673	1.51	1.11	.31	3745*	1.58	-.07	.20	3907	1.43	-1.08	.35
3804	.95	1.42	.35	3720*	1.45	.26	.29	3704*	1.39	-1.13	.23
3733*	1.24	1.40	.35	3607	1.38	.09	.35	3655	1.37	-.90	.35
3719*	1.18	1.08	.31	3811	1.15	.22	.35	3813	1.20	-.97	.17
3921*	.91	1.23	.29	3908	1.15	.07	.31	3919*	1.30	-.98	.21
3827	.87	1.35	.35	3649	1.32	.11	.22	3680	1.33	-1.01	.16
3716*	1.14	1.14	.27	3632	1.23	.27	.35	3808	.99	-1.00	.30
3642	1.11	1.11	.24	3718*	1.22	.16	.33	3686	1.26	-.88	.29
3902	.73	1.49	.29	3629	1.11	-.03	.35	3721*	1.23	-1.20	.22
3627	1.03	1.07	.35	3732*	.96	-.01	.35	3821	.90	-.92	.35
3681	1.03	1.34	.35	3633	.94	-.08	.35	3679	1.21	-.94	.17
3676	.89	1.51	.25	3609	.78	.18	.35	3685	1.19	-1.01	.16
3644	.88	1.25	.35	3730*	.75	.01	.10	3668	.97	-.87	.14
3717*	.83	1.25	.35	3618	.64	-.05	.00	3684	.86	-.85	.14
3670	.80	1.11	.35					3703*	.83	-1.16	.21
3647	.79	1.14	.35	Mean (F)	1.08	.09	.29	3617	.79	-1.11	.14
				Mean*(W)	1.17	.09	.28	3713*	.75	-1.18	.33
Mean (F)	1.05	1.26	.32								
Mean*(W)	1.05	1.25	.32					Mean (F)	1.19	-.97	.23
				Stratum 4 (19 items)				Mean*(W)	1.19	-1.01	.24
Stratum 7 (15 items)				3744*	1.94	-.35	.30				
3743*	2.14	.68	.32	3708*	1.62	-.20	.16	Stratum 1 (19 items)			
3661	1.90	.68	.32	3631	1.53	-.18	.35	3741*	1.63	-1.56	.35
3674	1.72	.63	.26	3814	1.26	-.32	.35	3910	1.58	-1.59	.21
3909	1.34	.77	.35	3903	1.21	-.43	.31	3692	1.53	-1.28	.35
3662	1.34	.93	.27	3671	1.31	-.14	.26	3825	1.09	-1.38	.34
3634	1.31	.84	.21	3701	.82	-.15	.35	3639	1.47	-1.80	.35
3669*	1.45	.70	.32	3643	1.40	-.30	.25	3638	1.35	-1.54	.21
3623*	1.42	.74	.31	3914	.98	-.39	.16	3913	1.31	-1.31	.19
3912	.95	.70	.19	3693	1.13	-.24	.24	3837*	1.09	-1.39	.25
3734*	.89	.96	.35	3725*	1.09	-.52	.24	3715*	1.16	-1.63	.26
3700	.84	.85	.30	3710*	1.02	-.33	.30	3920*	1.12	-1.34	.23
3659	1.37	.67	.29	3653	.83	-.51	.33	3842*	1.01	-1.33	.35
3635	1.17	.66	.35	3660	.78	-.39	.14	3695	1.09	-1.73	.22
3612	1.12	.75	.35	3922*	.64	-.26	.30	3731*	1.05	-1.67	.35
3616	.86	.62	.25	3606	.71	-.22	.14	3832	.99	-1.74	.32
				3663	.69	-.17	.33	3838*	.99	-1.68	.35
Mean (F)	1.32	.73	.29	3696	.68	-.35	.00	3613	.86	-1.74	.33
Mean*(W)	1.35	.75	.30	3656	.63	-.31	.34	3683	.85	-1.31	.14
								3657	.81	-1.74	.35
				Mean(F)	1.01	-.31	.25	3610	.80	-1.33	.14
				Mean*(W)	1.08	-.31	.26				
								Mean (F)	1.14	-1.34	.26
								Mean*(W)	1.15	-1.55	.28

Note. Items with asterisks are those which were added to the pool Winter quarter. All other items were in the pool both Fall and Winter quarters.

DO NOT FILM THIS PAGE

DISTRIBUTION LIST

Navy

1 Dr. Jack R. Borsting
Provost & Academic Dean
U.S. Naval Postgraduate School
Monterey, CA 93940

1 CCHNAVMILPERSCOM (N-6C)
Dept. of Navy
Washington, DC 20370

1 Dr. Richard Elster
Department of Administrative Sciences
Naval Postgraduate School
Monterey, CA 93940

1 DR. PAT FEDERICO
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Mr. Paul Foley
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. John Ford
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Patrick R. Harrison
Psychology Course Director
LEADERSHIP & LAW DEPT. (7b)
DIV. OF PROFESSIONAL DEVELOPMENT
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Dr. Norman J. Kerr
Chief of Naval Technical Training
Naval Air Station Memphis (75)
Millington, TN 38054

1 Dr. Leonard Kroeker
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. William L. Maloy
Principal Civilian Advisor for
Education and Training
Naval Training Command, Code 00A
Pensacola, FL 32508

1 Dr. Kneale Marshall
Scientific Advisor to DCNO(MPT)
OP01T
Washington DC 20370

1 CAPT Richard L. Martin, USN
Prospective Commanding Officer
USS Carl Vinson (CVN-70)
Newport News Shipbuilding and Drydock Co
Newport News, VA 23607

1 Dr. James McBride
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. George Moeller
Head, Human Factors Dept.
Naval Submarine Medical Research Lab
Groton, CN 06340

1 Dr William Montague
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. William Moonan
Code 303
Navy Personnel R&D Center
San Diego, CA 92152

1 LCDR W. Moroney
Code 55MP
Naval Postgraduate School
Monterey, CA 93940

1 Commanding Officer
U.S. Naval Amphibious School
Coronado, CA 92155

1 Library
Naval Health Research Center
P. O. Box 85122
San Diego, CA 92138

1 Ted M. I. Yellen
Technical Information Office, Code 201
NAVY PERSONNEL R&D CENTER
SAN DIEGO, CA 92152

1 Library, Code P201L
Navy Personnel R&D Center
San Diego, CA 92152

5 Technical Director
Navy Personnel R&D Center
San Diego, CA 92152

1 Director, Navy Personnel R&D Center
Washington Liason Office
Building 200, 2N
Washington Navy Yard, DC 20374

6 Commanding Officer
Naval Research Laboratory
Code 2627
Washington, DC 20390

1 Psychologist
ONR Branch Office
Bldg 114, Section D
666 Summer Street
Boston, MA 02210

1 Psychologist
ONR Branch Office
536 S. Clark Street
Chicago, IL 60605

1 Office of Naval Research
Code 437
800 N. Quincy SStreet
Arlington, VA 22217

5 Personnel & Training Research Programs
(Code 458)
Office of Naval Research
Arlington, VA 22217

1 Psychologist
ONR Branch Office
1030 East Green Street
Pasadena, CA 91101

1 Scientific Director
Office of Naval Research
Scientific Liaison Group/Tokyo
American Embassy
APO San Francisco, CA 96503

1 Special Asst. for Education and
Training (OP-01E)
Rm. 2705 Arlington Annex
Washington, DC 20370

1 Office of the Chief of Naval Operations
Research, Development, and Studies Branc
(OP-102)
Washington, DC 20350

1 Long-Range Manpower, Personnel, and
Training Planning Branch (OP-110)
Room 3828
Arlington Annex
Washington, DC 20350

1 Head, Manpower Training and Reserves
Section (Op-964D)
Room 4A478, The Pentagon
Washington, DC 20350

1 Captain Donald F. Parker, USN
Commanding Officer
Navy Personnel R&D Center
San Diego, CA 92152

1 LT Frank C. Petho, MSC, USN (Ph.D)
Code L51
Naval Aerospace Medical Research Laborat
Pensacola, FL 32508

1 DR. RICHARD A. POLLAK
ACADEMIC COMPUTING CENTER
U.S. NAVAL ACADEMY
ANNAPOLIS, MD 21402

1 Dr. Gary Poock
Operations Research Department
Code 55PK
Naval Postgraduate School
Monterey, CA 93940

1 The Principal Deputy Assistant
Secretary of the Navy (MRA&L)
4E780, The Pentagon
Washington, DC 20350

1 Dr. Richard Sorensen
Navy Personnel R&D Center
San Diego, CA 92152

1 Dr. Ronald Weitzman
Code 54 WZ
Department of Administrative Sciences
U. S. Naval Postgraduate School
Monterey, CA 93940

1 DR. H.M. WEST III (OP-01)
Head, Program Development Branch(OP-120)
ARLINGTON ANNEX
WASHINGTON, DC 20350

1 Dr. Robert Wisher
Code 309
Navy Personnel R&D Center
San Diego, CA 92152

1 DR. MARTIN F. WISKOFF
NAVY PERSONNEL R& D CENTER
SAN DIEGO, CA 92152

1 Director, Research & Analysis Division
Plans and Policy Department
Navy Recruiting Command
4015 Wilson Boulevard
Arlington, VA 22203

Previous Publications

Proceedings of the 1977 Computerized Adaptive Testing Conference. July 1978.

Research Reports

79-7. The Person Response Curve: Fit of Individuals to Item Characteristic Curve Models. December 1979.
79-6. Efficiency of an Adaptive Inter-Subtest Branching Strategy in the Measurement of Classroom Achievement. November 1979.
79-5. An Adaptive Testing Strategy for Mastery Decisions. September 1979.
79-4. Effect of Point-in-Time in Instruction on the Measurement of Achievement. August 1979.
79-3. Relationships among Achievement Level Estimates from Three Item Characteristic Curve Scoring Methods. April 1979.
Final Report: Bias-Free Computerized Testing. March 1979.
79-2. Effects of Computerized Adaptive Testing on Black and White Students. March 1979.
79-1. Computer Programs for Scoring Test Data with Item Characteristic Curve Models. February 1979.
78-5. An Item Bias Investigation of a Standardized Aptitude Test. December 1978.
78-4. A Construct Validation of Adaptive Achievement Testing. November 1978.
78-3. A Comparison of Levels and Dimensions of Performance in Black and White Groups on Tests of Vocabulary, Mathematics, and Spatial Ability. October 1978.
78-2. The Effects of Knowledge of Results and Test Difficulty on Ability Test Performance and Psychological Reactions to Testing. September 1978.
78-1. A Comparison of the Fairness of Adaptive and Conventional Testing Strategies. August 1978.
77-7. An Information Comparison of Conventional and Adaptive Tests in the Measurement of Classroom Achievement. October 1977.
77-6. An Adaptive Testing Strategy for Achievement Test Batteries. October 1977.
77-5. Calibration of an Item Pool for the Adaptive Measurement of Achievement. September 1977.
77-4. A Rapid Item-Search Procedure for Bayesian Adaptive Testing. May 1977.
77-3. Accuracy of Perceived Test-Item Difficulties. May 1977.
77-2. A Comparison of Information Functions of Multiple-Choice and Free-Response Vocabulary Items. April 1977.
77-1. Applications of Computerized Adaptive Testing. March 1977.
Final Report: Computerized Ability Testing, 1972-1975. April 1976.
76-5. Effects of Item Characteristics on Test Fairness. December 1976.
76-4. Psychological Effects of Immediate Knowledge of Results and Adaptive Ability Testing. June 1976.
76-3. Effects of Immediate Knowledge of Results and Adaptive Testing on Ability Test Performance. June 1976.
76-2. Effects of Time Limits on Test-Taking Behavior. April 1976.
76-1. Some Properties of a Bayesian Adaptive Ability Testing Strategy. March 1976.
75-6. A Simulation Study of Stradaptive Ability Testing. December 1975.
75-5. Computerized Adaptive Trait Measurement: Problems and Prospects. November 1975.
75-4. A Study of Computer-Administered Stradaptive Ability Testing. October 1975.
75-3. Empirical and Simulation Studies of Flexilevel Ability Testing. July 1975.
75-2. TETREST: A FORTRAN IV Program for Calculating Tetrachoric Correlations. March 1975.
75-1. An Empirical Comparison of Two-Stage and Pyramidal Adaptive Ability Testing. February 1975.
74-5. Strategies of Adaptive Ability Measurement. December 1974.
74-4. Simulation Studies of Two-Stage Ability Testing. October 1974.
74-3. An Empirical Investigation of Computer-Administered Pyramidal Ability Testing. July 1974.
74-2. A Word Knowledge Item Pool for Adaptive Ability Measurement. June 1974.
74-1. A Computer Software System for Adaptive Ability Measurement. January 1974.
73-3. The Stratified Adaptive Computerized Ability Test. September 1973.
73-2. Comparison of Four Empirical Item Scoring Procedures. August 1973.
73-1. Ability Measurement: Conventional or Adaptive? February 1973.

Copies of these reports are available, while supplies last, from:
Computerized Adaptive Testing Laboratory
Psychometric Methods Program, Department of Psychology
N660 Elliott Hall, University of Minnesota
75 East River Road, Minneapolis, Minnesota 55455